

systems biology research group



ALE Analytics - A microbial data warehouse and analytical suite

Patrick V. Phaneuf¹, Dennis Gosting², Bernhard O. Palsson^{2,3}, Adam M. Feist^{2,3}

¹) Department of Computer Science and Engineering, University of California at San Diego, La Jolla 92093-0404

²) Department of Bioengineering, University of California San Diego, La Jolla, CA, 92093, United States of America,

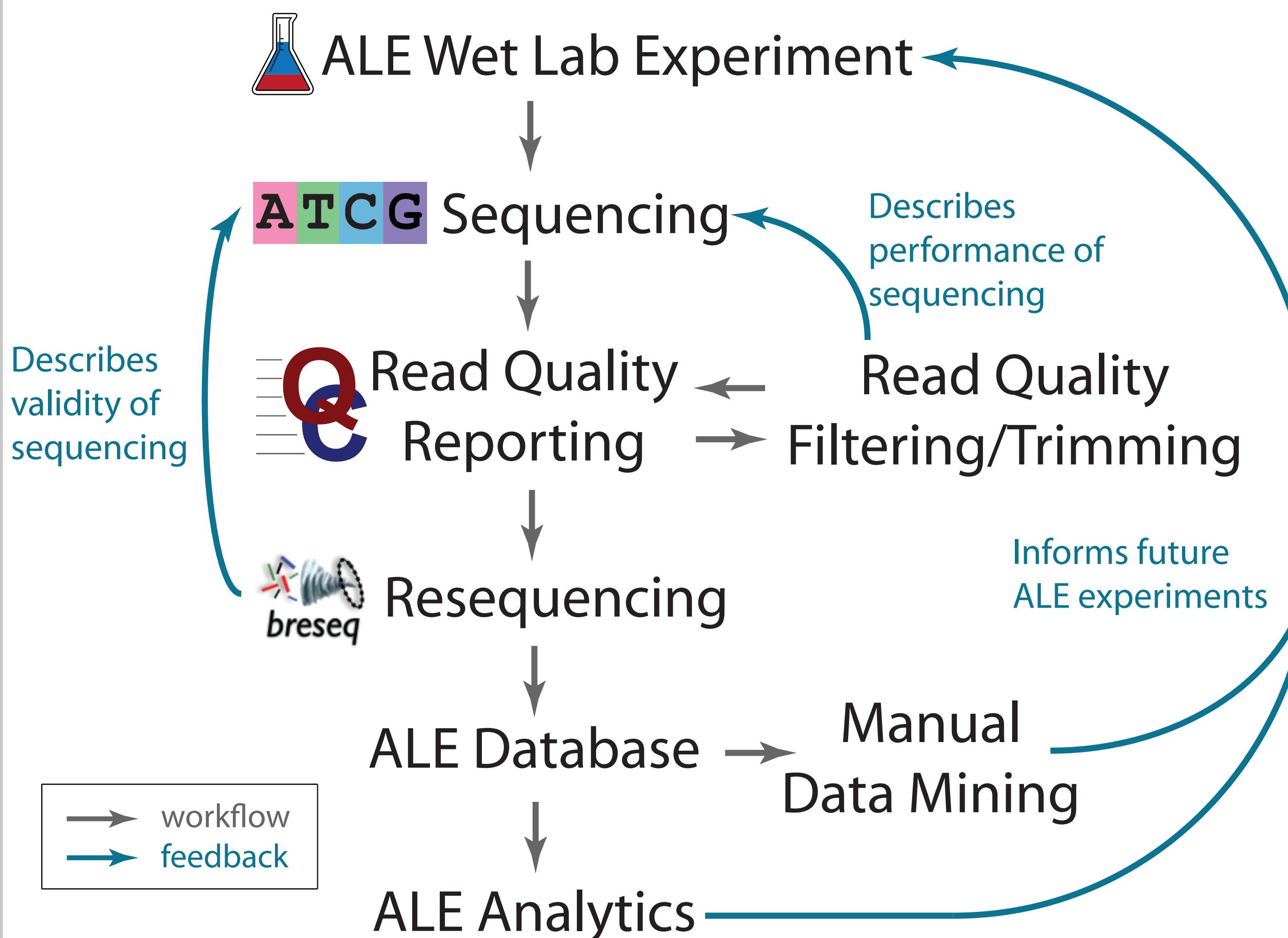
³) Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, 2800, Lyngby, Denmark

Abstract

Adaptive Laboratory Evolution (ALE) is a powerful experimental tool used by academic and industrial labs for a number of applications. ALE experiments generate a substantial amount of data, coming in the form of sequencing reads, alignment reports, and sample metadata. As ALE experiments scale to include more samples, the task of managing this data comes at higher costs due to the effort necessary to organize and integrate data into a format that describes the evolution process succinctly. In this work, we describe the development, deployment and iteration of an 'ALE Analytics' pipeline and web platform that streamlines the necessary ALE experiment data post-processing, manages experiment data, and produces interactive reports that detail an ALE experiment. Our design has been primarily driven by the need to consolidate large amounts of ALE experimental data in such a way to describe the quality of the sample sequencing, adaptive mutations in evolved strains, the context of mutations via their metadata (i.e., culturing environments, strain properties), and related mutations found in other experiments housed in the database. We have done so by leveraging a full stack of technologies that enable the parsing and databasing of experiment data, the execution of automated analysis on said data and the generation of web accessible reports. Future efforts will take full advantage of this developing platform to enable more depth and breadth of ALE experiment analysis with quicker turnaround.

ALE Mutation Processing Pipeline

Our ALE pipeline leverages industry established tools to process and refine ALE replicate sequencing data to identify genomic perturbations and integrate them into a database that can be mined for mutational trends.



Features and Advantages

- Supports multiple bacterial strains.
- Leverages established tools: Breseq, FastQC, FASTX-Toolkit, etc.
- Automated multi-sample processing for specific stages.
- Stage dependent feedback.
- Feedback provides opportunity for experiment and data refinement.

ALE Analytics Web Platform

The ALE Analytics web platform integrates our ALE mutation analysis with published multi-omics data and better enables identification of causal mutations in ALE experiments.

Dashboard

General statistics on entire ALE Database including:

- Most frequently mutated genes
- Most frequent mutations
- Mutation type counts (SNP, INDEL, etc.)

| Mutation Type | Count |
|-----------------------------|-------|
| Single Base Substitutions | 11163 |
| Multiple Base Substitutions | 41 |
| Deletions | 959 |
| Insertions | 377 |
| Mobile Element Insertions | 137 |
| Amplifications | 3 |
| Gene Conversions | 0 |
| Inversions | 0 |
| Duplications | 1052 |

| Functional Change Type | Count |
|------------------------|-------|
| Intergenic | 3564 |
| Noncoding | 351 |
| Pseudogenes | 337 |
| Synonymous | 2044 |
| Nonsynonymous | 5090 |

Top Genes:

- rsH (251)
- rsH (1652)
- rsA (1145)
- carB (1000)
- gtpPyG (559)

Genes

Gene specific information:

- All mutations in the ALE Database involving a particular gene.
- 3D rendering of a gene's product.
- Highlight structural area of a gene's product effected by a mutation.
- Displays ligand structure at site of mutation.

Mutation Search

Search the ALE Database by mutation specific features.

References

Original paper and those I report biological findings on. Additionally, any 3rd party software I mention.

Contact and Info

pphaneuf@eng.ucsd.edu
afeist@ucsd.edu
systemsbiology.ucsd.edu

Future Directions

- Integrate ALE experiment fitness data to provide additional context for identifying causal mutations.
- Use whole ALE Database statistics to rank the novelty of causal mutations.
- Datamine for pairwise mutations and establish their relationship.
- Datamine for novel regulatory genomic regions.